

Input file Fbh55053e.seq; Output File 55053.trans  
Sequence length 2862

GTCGACCCACGCGTCCGCGGACGCGTGGGCGGACGCGTGGGCGG

ACGCGTCCGGGGGGACCGGTCCGGGCCGGGACCAAGGGCACC	M	S	S	G	A	K	E	G	G	9
ATG TCG TCC GGG GCC AAG GAG GGA GGT	27									
G G S P A Y H L P H P H P H P P Q H A Q	29									
GGG GGC TCT CCC GCC TAC CAC CTC CCT CAC CCC CAC CCC CAC CCA CCC CAG CAC GCC CAA	87									
Y V G P Y R L E K T L G K G Q T G L V K	49									
TAT GTG GGC CCC TAT CGG CTG GAG AAG ACG CTG GGC AAA GGA CAG ACA GGG CTG GTT AAA	147									
L G V H C I T G Q K V A I K I V N R E K	69									
CTC GGG GTC CAC TGC ATC ACG GGT CAG AAG GTC GCC ATC AAG ATC GTG AAC CGG GAG AAG	207									
L S E S V L M K V E R E I A I L K L I E	89									
CTG TCG GAG TCG GTG CTG ATG AAG GTG GAG CGG GAG ATC GCC ATC CTG AAG CTC ATC GAA	267									
H P H V L K L H D V Y E N K K Y L Y L V	109									
CAC CCA CAT GTC CTC AAG CTC CAC GAC GTC TAC GAG AAC AAG AAA TAT TTG TAC CTG GTT	327									
L E H V S G G E L F D Y L V K K G R L T	129									
CTG GAG CAC GTC TCG GGG GGT GAG CTA TTC GAC TAC CTG GTA AAG AAG GGG AGA CTG ACG	387									
P K E A R K F F R Q I V S A L D F C H S	149									
CCC AAG GAG GCC CGA AAG TTC TTC CGC CAG ATT GTG TCT GCG CTG GAC TTC TGC CAC AGC	447									
Y S I C H R D L K P E N L L L D E K N N	169									
TAC TCC ATC TGC CAC AGA GAC CTA AAG CCC GAG AAC CTG CTT TTG GAT GAG AAA AAC AAC	507									
I R I A D F G M A S L Q V G D S L L E T	189									
ATC CGC ATT GCA GAC TTC GGC ATG GCG TCC CTG CAG GTG GGG GAC AGC CTC CTG GAG ACC	567									
S C G S P H Y A C P E V I K G E K Y D G	209									
AGC TGC GGG TCC CCC CAT TAT GCG TGT CCA GAG GTG ATT AAG GGG GAA AAA TAT GAT GGC	627									
R R A D M W S C G V I L F A L L V G A L	229									
CGC CGG GCA GAC ATG TGG AGC TGT GGA GTC ATC CTC TTC GCC CTG CTC GTG GGG GCT CTG	687									
P F D D D N L R Q L L E K V K R G V F H	249									
CCC TTT GAT GAC GAC AAC CTC CGC CAG CTG CTG GAG AAG GTG AAA CGG GGC GTC TTC CAC	747									
M P H F I P P D C Q S L L R G M I E V E	269									
ATG CCC CAC TTC ATT CCT CCA GAT TGC CAG AGC CTC CTG AGG GGA ATG ATC GAA GTG GAG	807									
P E K R L S L E Q I Q K H P W Y L G G K	289									
CCC GAA AAA AGG CTC AGT CTG GAG CAA ATT CAG AAA CAT CCT TGG TAC CTA GGC GGG AAA	867									
H E P D P C L E P A P G R R V A M R S L	309									
CAC GAG CCA GAC CCG TGC CTG GAG CCA GCC CCT GGC CGC CGG GTA GCC ATG CGG AGC CTG	927									
P S N G E L D P D V L E S M A S L G C F	329									
CCA TCC AAC GGA GAG CTG GAC CCC GAC GTC CTA GAG AGC ATG GCA TCA CTG GGC TGC TTC	987									
R D R E R L H R E L R S E E E N Q E K M	349									
AGG GAC CGC GAG AGG CTG CAT CGC GAG CTG CGC AGT GAG GAG GAG AAC CAA GAA AAG ATG	1047									
I Y Y L L L D R K E R Y P S C E D Q D L	369									
ATA TAT TAT CTG CTT TTG GAT CGG AAG GAG CGG TAT CCC AGC TGT GAG GAC CAG GAC CTG	1107									
P P R N D V D P P R K R V D S P M L S R	389									
CCT CCC CGG AAT GAT GTT GAC CCC CCC CGG AAG CGT GTG GAT TCT CCC ATG CTG AGC CGT	1167									
H G K R R P E R K S M E V L S I T D A G	409									
CAC GGG AAG CGG CGA CCA GAG CGG AAG TCC ATG GAA GTC CTG AGC ATC ACC GAT GCC GGG	1227									
G G G S P V P T R R A L E M A Q H S Q R	429									
GGT GGT GGC TCC CCT GTA CCC ACC CGA CGG GCC TTG GAG ATG GCC CAG CAC AGC CAG AGA	1287									
S R S V S G A S T G L S S S P L S S P R	449									
TCC CGT AGC GTC AGT GGA GCC TCC ACG GGT CTG TCC TCC AGC CCT CTA AGC AGC CCA AGG	1347									

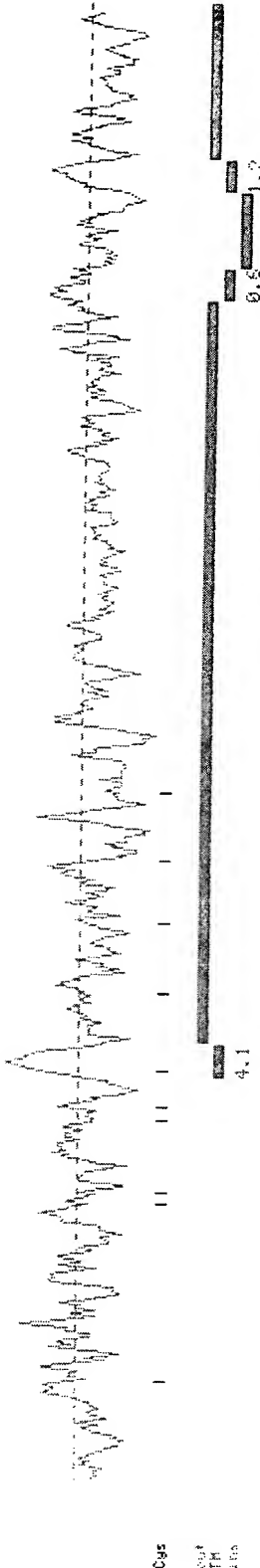
Fig. 1A

S	P	V	F	S	F	S	P	E	P	G	A	G	D	E	A	R	G	G	G	469
AGT	CCG	GTC	TTT	TCC	TTT	TCA	CCG	GAG	CCG	GGG	GCT	GGA	GAT	GAG	GCT	CGA	GGC	GGG	GGC	1407
S	P	T	S	K	T	Q	T	L	P	S	R	G	P	R	G	G	G	A	G	489
TCC	CCG	ACT	TCC	AAA	ACG	CAG	ACG	CTG	CCT	TCT	CGG	GGC	CCC	AGG	GGT	GGG	GGC	GCC	GGG	1467
E	Q	P	P	P	P	S	A	R	S	T	P	L	P	G	P	P	G	S	P	509
GAG	CAG	CCC	CCG	CCC	CCC	AGT	GCC	CGC	TCC	ACA	CCC	CTG	CCC	GGC	CCC	CCA	GGC	TCC	CCG	1527
R	S	S	G	G	T	P	L	H	S	P	L	H	T	P	R	A	S	P	T	529
CGC	TCC	TCT	GGC	GGG	ACC	CCC	TTG	CAC	TCG	CCT	CTG	CAC	ACG	CCC	CGG	GCC	AGT	CCC	ACC	1587
G	T	P	G	T	T	P	P	P	S	P	G	G	G	V	G	G	A	A	W	549
GGG	ACC	CCG	GGG	ACA	ACA	CCA	CCC	CCC	AGC	CCC	GGC	GGT	GGC	GTC	GGG	GGG	GCC	GCC	TGG	1647
R	S	R	L	N	S	I	R	N	S	F	L	G	S	P	R	F	H	R	R	569
AGG	AGT	CGT	CTC	AAC	TCC	ATC	CGC	AAC	AGC	TTC	CTG	GGC	TCC	CCT	CGC	TTT	CAC	CGG	CGC	1707
K	M	Q	V	P	T	A	E	E	M	S	S	L	T	P	E	S	S	P	E	589
AAG	ATG	CAG	GTC	CCT	ACC	GCT	GAG	GAG	ATG	TCC	AGC	TTG	ACG	CCA	GAG	TCC	TCC	CCG	GAG	1767
L	A	K	R	S	W	F	G	N	F	I	S	L	D	K	E	E	Q	I	F	609
CTG	GCA	AAA	CGC	TCC	TGG	TTC	GGG	AAC	TTC	ATC	TCC	TTG	GAC	AAA	GAA	GAA	CAA	ATA	TTC	1827
L	V	L	K	D	K	P	L	S	S	I	K	A	D	I	V	H	A	F	L	629
CTC	GTG	CTA	AAG	GAC	AAA	CCT	CTC	AGC	AGC	ATC	AAA	GCA	GAC	ATC	GTC	CAT	GCC	TTT	CTG	1887
S	I	P	S	L	S	H	S	V	L	S	Q	T	S	F	R	A	E	Y	K	649
TCG	ATC	CCC	AGC	CTG	AGT	CAC	AGT	GTG	CTG	TCA	CAG	ACC	AGC	TTC	AGG	GCC	GAG	TAC	AAG	1947
A	S	G	G	P	S	V	F	Q	K	P	V	R	F	Q	V	D	I	S	S	669
GCC	AGT	GGC	GGC	CCC	TCC	GTC	TTC	CAA	AAG	CCC	GTC	CGC	TTC	CAG	GTG	GAC	ATC	AGC	TCC	2007
S	E	G	P	E	P	S	P	R	R	D	G	S	G	G	G	G	I	Y	S	689
TCT	GAG	GGT	CCA	GAG	CCC	TCC	CCG	CGA	CGG	GAC	GGC	AGC	GGA	GGT	GGT	GGC	ATC	TAC	TCC	2067
V	T	F	T	L	I	S	G	P	S	R	R	F	K	R	V	V	E	T	I	709
GTC	ACC	TTC	ACT	CTC	ATC	TCG	GGT	CCC	AGC	CGT	CGG	TTC	AAG	CGA	GTG	GTG	GAG	ACC	ATC	2127
Q	A	Q	L	L	S	T	H	D	Q	P	S	V	Q	A	L	A	D	E	K	729
CAG	GCA	CAG	CTC	CTG	AGC	ACT	CAT	GAC	CAG	CCC	TCC	GTG	CAG	GCC	CTG	GCA	GAC	GAG	AAG	2187
N	G	A	Q	T	R	P	A	G	A	P	P	R	S	L	Q	P	P	P	G	749
AAC	GGG	GCC	CAG	ACC	CGG	CCT	GCT	GGT	GCC	CCA	CCC	CGA	AGC	CTG	CAG	CCC	CCA	CCC	GGC	2247
R	P	D	P	E	L	S	S	S	P	R	R	G	P	P	K	D	K	K	L	769
CGC	CCA	GAC	CCA	GAG	CTG	AGC	AGC	TCT	CCC	CGC	CGA	GGC	CCC	CCC	AAG	GAC	AAG	AAG	CTC	2307
L	A	T	N	G	T	P	L	P	*											779
CTG	GCC	ACC	AAC	GGG	ACC	CCT	CTG	CCC	TGA											2337

CCCCACGGGGCCGGGGAGGGAGGGGACCCCCCTCCACCCCCCTTCCGTGCCCCCAACTGTGAATCTGTAAATAAGGCC  
 CAAGGAACATGTCTGGGAGGGGGTGGACACAAAAACCGGCCTTGCCCTGCAGGGATGGGGCTCCACAGGCCGTGCCAA  
 CTGGGGGTGGTTCTAGGGGAACAGGGGGCGGGGAGCTGTTTCTATTTTATTTATTGATTAATTTATTTATTTATTTAT  
 TGATCAATCTCTCTGCGGGGTGGGGTGGGGGAGGGACGGGAGCTGGTTGGGGTGGCTTAGCAGATCCGGACAGGGCCCT  
 CTGTCCCTGTGTCTGTCCTCCCAACCCCCCTCTCCCGGGCCCCCTCCCTCCCTGGTCCCTCCCCCACGACCTTCTGTACGGAT  
 TTGCTCTCCGGAAGGAATTCTGGTTTCGCGTGATCCTGCCTGCGTCCGTGTCTCTGATTCCGCCGGCGGCAAAAAAAAAA  
 AAAAAA: AAAAAAAAAAAAAAAAAAAGATAATAATAATAAATAGCCTTGATCAGGGAA

Fig. 1B

# Analysis of 55053 (778 aa)



>55053  
 MSSGAKEGGGSPAYHLPHPHPQHAQYVGPYRLKLTGKGOTGLVKLVGHCHITGQKV  
 AIKIVNREKLSVLMKVEREIAILKLIHPHVLKLDVYENKYLVLVHVSGGELFD  
 YLVKGRLLTPKEARKFFRQIVSALDFCHSYISICHRDLKPENLLDEKNNIRIADFGMASL  
 QVGDLSLETSCGPHYACPEVIKGEKYDGRADMMSCGVILFALLVGLPFDDNLRQLL  
 EKVRGVFIHPHFIPDCQSLLRMIIEVEPEKRLSLEQIQKHFWYLGKHEPDPCLEPAP  
 GRRVAMRSLPSNGELDFDVLSEHSLGCFDRERLHRELSEENQEKMIYYILLDRKER  
 YPSCEDQDLPPRNDVDPKRKVDSPMLSRHGKRRPERKSMENVLSITDAGGGSPVPTTRA  
 LEMAHOSRORSVSGASTGLSSPLSSPSRVFSEFPGAGDEARGGSPSTSKTQTLPS  
 RGPFGGAGEQPPPSARSTPLPGPSRSGGTPLHSPHTPRASPTGTPTTTPPSP  
 GGVGGAARSRNLNIRNSFLGSPFRHKKQVPTAEEMSSLTPESSPELAKRSWFGNFI  
 SLDKKEQIFVLKDKPLSSIKADIVHAFLSIPSLSHSVLSOTSFRAEYKASGSPSVTKP  
 VRFQVDISSESEPEPSRRDGGGGIYSVTFTLISGFSRRFRKVVETIQALLSTHDQP  
 SVQALADEKNGAQRAGAPRSIQPPGRDPDELSSSPRRGPPKDKLLATNGTPLP

Transmembrane Segments Predicted by MEMSAT

Start	End	Orient	Score
214	231	ins->out	4.1
624	640	out->ins	0.6
681	697	ins->out	1.2

Fig. 2

## Protein Family / Domain Matches, HMMer version 2

Searching for complete domains in PFAM  
 hmmpfam - search a single seq against HMM database  
 HMMER 2.1.1 (Dec 1998)  
 Copyright (C) 1992-1998 Washington University School of Medicine  
 HMMER is freely distributed under the GNU General Public License (GPL).

HMM file: /prod/ddm/seqanal/PFAM/pfam5.5/Pfam  
 Sequence file: /prod/ddm/wspace/orfanal/oa-script.23506.seq  
 Query: 55053

Scores for sequence family classification (score includes all domains):

Model	Description	Score	E-value	N
pkinase	Eukaryotic protein kinase domain	323.4	2.6e-93	1
UBA	UBA domain	7.7	4.9	1

Parsed for domains:

Model	Domain	seq-f	seq-t	hmm-f	hmm-t	score	E-value
pkinase	1/1	34	285	1	278	323.4	2.6e-93
UBA	1/1	315	356	1	41	7.7	4.9

Alignments of top-scoring domains:

pkinase: domain 1 of 1, from 34 to 285: score 323.4, E = 2.6e-93

```

      *->yelleklGeGsfGkVykakhk.tgkivAvKilkkelsl.....lr
      y+l ++lG+G G V++++h tg++vA+Ki+++e+ls++ + r
55053 34 YRLEKTLGKGQTGLVKLGVHCITGQKVAIKIVNREKLSesvlmkvER 80

      EiqilkrslHpnIvrlgvfedtdhlylvmEymegGdLfdylrrngpls
      Ei+ilk + Hp++++l++v+e +++lylv+E++ gG+Lfdyl+++g+l+
55053 81 EIAILKLIIEHPVHLKLHDVYE-NKKYLYLVLEHVSGGELFDYLVKKGRLT 129

      ekeakkialQilrGleYlHsngivHRDLKpeNlLldengtvKiaDFGLAr
      +kea+k+++Qi+++l+++Hs +i+HRDLKpeN+Llde+++++iaDFG+A
55053 130 PKEARKFFRQIVSALDFCHSYISICHRDLKPENLLLEKNNIRIADFGMAS 179

      ll...eklttfvGtpwYmmAPEvileg.rgysskvDvWSlGvilyElltg
      l +++ l t +G+p+Y PEv +g++++++D+WS+GvIL+ ll g
55053 180 LQvgdSLLTSCGSPHYA-CPEV-IKGeKYDGRADMWSCGVILFALLVG 227

      gplfpgadlpaftggdevdqliifvklPfsdelpktridpleelfrik
      lPf+d d+l+l++ +k
55053 228 -----ALPFDD-----DNLRLQLEKVK 244

      r.rlpplsncSeelkdLlkkcLnkDPskRpGsatakeilnhpwf<-*
      r+ + p+ ++++++Ll+++++ P+kR+ + ++i +hpw
55053 245 RgVFHMPHFIPDCQSLLRGMIEVEPEKRL---SLEQIQKHPWY 285

UBA: domain 1 of 1, from 315 to 356: score 7.7, E = 4.9
      *->edeekieqlvEMGF..dreevvkALratngngverAaewLlsh<-*
      d + +e+ +G +dre+ + Lr+ n e+ +++Ll +
55053 315 LDPDVLESMAISLGCfrDRERLHRELRLSEEN-QEKMIYYLLLD 356
  
```

//  
 Searching for complete domains in SMART  
 hmmpfam - search a single seq against HMM database  
 HMMER 2.1.1 (Dec 1998)  
 Copyright (C) 1992-1998 Washington University School of Medicine  
 HMMER is freely distributed under the GNU General Public License (GPL).

HMM file: /ddm/robison/smart/smart/smart.all.hmms  
 Sequence file: /prod/ddm/wspace/orfanal/oa-script.23506.seq  
 Query: 55053

Fig. 3A

Scores for sequence family classification (score includes all domains):

Model	Description	Score	E-value	N
serkin_6		356.8	2.4e-103	1
tyrkin_6		39.2	2.4e-14	1

Parsed for domains:

Model	Domain	seq-f	seq-t	hmm-f	hmm-t	score	E-value
serkin_6	1/1	34	285	1	231	356.8	2.4e-103
tyrkin_6	1/1	34	286	1	280	39.2	2.4e-14

Alignments of top-scoring domains:

serkin\_6: domain 1 of 1, from 34 to 285: score 356.8, E = 2.4e-103

```

*->YellkklGkGaFGkVylardkktgrlvAiKvik.....erilr
Y+l k+lGkG G V+l+++ tg++vAiK+++++ +++ +rr
55053 34 YRLEKTlGKGQTGLVKLGvHCITGQKVAIKlVNreklseSVlMKVER 80

EikiLkk.dHPNIVKlydvfed.dklylVmEyceGdlGdLfdllkkrgr
Ei+iLk HP++ kL+dv+e++++lylV+E+++G G+Lfd+l+k+gr
55053 81 EIAiLKLIEHPHVLKLDVYENkKYLYLVLEHVSG--GELFDYLvKKGR- 127

glrkvlseE.earfyfrQilsaLeYLHsqqIiHRDLKPeNiLLds..hvKl
l+++ear++frQi+saL+++Hs I+HRDLKPeN+LLd+++++
55053 128 -----LTPKEARKFFRQIVSALDFCHSYSICHRLKPeNLLLDEKknNIRI 172

aDFGlarql.....ttfvGTpeYmAPEvl...gYgkpavDiWSlGcily
aDFG+A + +++ t +G+p+Y PEv+++++Y+++++D+WS+G+il+
55053 173 ADFGMASLQvgdsllETSCGSPHYACPEVIKgeKYDGRRADMWSCGVILF 222

ElltGkpPFp..qldlifkkg.....SpeakdLikklLvkdPek
ll+G PF++++l +++k++++ + ++ *p++ +L++++ +++Pek
55053 223 ALLVGALPFDDdNLRQLLEKVKrgvfhmphfiPPDCQSLLRGMIEVEPEK 272

Rlta.eaLedeldikaHPff<--*
Rl+ +++ + HP+
55053 273 RLSLeQIQK-----HPWY 285

```

Fig. 3B

tyrkin\_6: domain 1 of 1, from 34 to 286: score 39.2, E = 2.4e-14  
 \*->ltlgkkLGeGaFGeVykGtlk...ieVAVKtLkeda....keeFlr  
 +l+k+LG G+ G V +G+ +++++VA+K ++ ++ ++ r  
 55053 34 YRLEKTLGKGQTGLVKLGVHCitgQKVAIKIVNREKlsesvLMKVER 80  
 EakiMkklGgkHpNiVklLlGvcteeerrFmevePlmivmEymegGdLldy  
 E+ i+k + +Hp+++kL+ v + + l++v+E+++gG L dy  
 55053 81 EIAILKLI--EHPHVLKLHDVYENK-----KYLVLVLEHVSGGELFDY 121  
 LrknrpklslsdLlsfAlQIAkGMeYLesknfvHRDLAARNcLvgenkvv  
 L k+++ l++++ +f QI + ++ +s + HRDL N L+++ +  
 55053 122 LVKKGR-LTPKEARKFFRQIVSALDFCHSYSICHRDLKPENLLLDEKNNI 170  
 KIsDFGLsRdlyddDkkGeskdYrkkgggkgtllPirwmAPESlkdgk  
 +I+DFG++ d + ++ g+ PE++k k  
 55053 171 RIADFGMASLQVGD-----SLLETSC--GSP-----HYACPEVIKGEK 206  
 Ft.skSDVWSFGVILWEiftlGeqPYpgeiqqfmsneevleylkkGyRlp  
 + + D WS GV L+ ++ G+ P + + +++le++k+G  
 55053 207 YDgRRADMWSCGVILFALL-VGALPFDD-----DNLRLLEKVKRG-VFH 249  
 kPendlpisSvtCPdelydIMlqCwaedPedRPtFsel...verl<-\*  
 P+ P++ +l + + +Pe+R + +++++ +l  
 55053 250 MPHf-----IPDCQSLLRGMI EVEPEKRLSLEQIqkhPWYL 286

//

Fig. 3C